The Lung Image Database Consortium (LIDC)

Published in final edited form as:
Acad Radiol. 2007 December ; 14(12): 1475–1485.

The Lung Image Database Consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements

Anthony P. Reeves[a], **Alberto M. Biancardi**[a], **Tatiyana V. Apanasovich**[b], **Charles R. Meyer**[c], **Heber MacMahon**[d], **Edwin J.R. van Beek**[e], **Ella A. Kazerooni**[c], **David Yankelevitz**[f], **Michael F. McNitt-Gray**[g], **Geoffrey McLennan**[h], **Samuel G. Armato III**[d], **Claudia I. Henschke**[f], **Denise R. Aberle**[g], **Barbara Y. Croft**[i], and **Laurence P. Clarke**[i]

a *School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA*

b *Operational Research and Information Engineering, Cornell University, Ithaca, NY, USA*

c *Department of Radiology, The University of Michigan, Ann Arbor, MI, USA*

d *Department of Radiology, The University of Chicago, Chicago, IL, USA*

e *Department of Radiology, University of Iowa, Iowa City, IA, USA*

f *Weill Medical College, Cornell University, New York, NY, USA*

g *David Geffen School of Medicine at UCLA, Los Angeles, CA, USA*

h *Medicine and Biomedical Engr., University of Iowa, Iowa City, IA, USA*

i *Cancer Imaging Program, National Cancer Institute, Bethesda, MD, USA*

Abstract

Rationale and Objectives—To investigate the effects of choosing between different metrics in estimating the size of pulmonary nodules as a factor both of nodule characterization and of performance of computer aided detection systems, since the latters are always qualified with respect to a given size range of nodules.

Materials and Methods—This study used 265 whole-lung CT scans documented by the Lung Image Database Consortium using their protocol for nodule evaluation. Each inspected lesion was reviewed independently by four experienced radiologists who provided boundary markings for nodules larger than 3 mm. Four size metrics, based on the boundary markings, were considered: a uni-dimensional and two bi-dimensional measures on a single image slice and a volumetric measurement based on all the image slices. The radiologist boundaries were processed and those with four markings were analyzed to characterize the inter-radiologist variation, while those with at least one marking were used to examine the difference between the metrics.

Results—The processing of the annotations found 127 nodules marked by all of the four radiologists and an extended set of 518 nodules each having at least one observation with three-dimensional sizes ranging from 2.03 to 29.4 mm (average 7.05 mm, median 5.71 mm). A very high inter-observer variation was observed for all these metrics: 95% of estimated standard deviations were in the following ranges [0.49, 1.25], [0.67, 2.55], [0.78, 2.11], and [0.96, 2.69] for the three-dimensional,

Corresponding Author: Anthony P. Reeves, Mail Address: Electrical and Computer Engr., Rhodes Hall 392, Cornell University, Ithaca, NY 14853, U.S.A., E-mail: reeves@ece.cornell.edu, Phone: +1.607.255.2342, Fax: +1.607.255.9072.

the uni-dimensional, and the two bi-dimensional size metrics respectively (in mm). Also a very large difference among the metrics was observed: 0.95 probability-coverage region widths for the volume estimation conditional on uni-dimensional, and the two bi-dimensional size measurements of 10mm were 7.32, 7.72, and 6.29 mm respectively.

Conclusions—The selection of data subsets for performance evaluation is highly impacted by the size metric choice. The LIDC plans to include a single size measure for each nodule in its database. This metric is not intended as a gold standard for nodule size; rather, it is intended to facilitate the selection of unique repeatable size limited nodule subsets.

Keywords

Quantitative image analysis; X-ray CT; Detection; Lung nodule annotation; Size metrics

1 Introduction

Accurate and reliable measurement of pulmonary nodule size from CT scans has an important role in computer assisted evaluation of lung lesions. It is a key factor in the diagnosis of lung cancer as the estimation of nodule growth rates serves as a predictor of malignancy; size change can also be used to assess the efficacy of a therapeutic treatment. Additionally, nodule sizing is a critical aspect of computer assisted diagnosis (CAD) systems, and in particular their detection sub-systems, because they are always qualified with respect to a given size range of nodules. The usual approach is to characterize a system based on its performance on a subset of images from a documented image database having a specified size range. In the context of spatial extent for a three-dimensional object without restrictions on shape, the size is best expressed by the volume occupied by that object. However, other important considerations in choosing a method for estimation of size also include the imaging modality and the time available to the physician. Image modalities may be only two-dimensional or highly anisotropic with regards to the third dimension. Manually measuring the lesion volume involves inspecting all images that include the lesion — a process that is very time consuming. To provide a standard method for lesion size measurement, the World Health Organization (WHO) proposed in 1979 the use of the product of the maximal diameter and its largest perpendicular [1], while the Response Evaluation Criteria in Solid Tumors (RECIST) working group in 1998–2000 proposed the use of the (uni-dimensional) maximal diameter as a more efficient standard estimator of lesion volume [2]. Another measure that has been seen in some present studies, [3,4], is the bi-dimensional Modified Schwartz equation (MS), first introduced by Usuda et al. in 1994 [5]. This measure is similar to the WHO measure in using the maximal diameter and its largest perpendicular, but differs from WHO because the lesion is assumed to be an ellipsoid with a long axis equal to the lesion maximal diameter and with equal-length short axes equal to the largest perpendicular.

Addressing the problem from a different perspective is the active interest in developing computer assisted methods that will aid the physician in measuring the size of lesions using volumetric methods [6–11]. The challenge here is how to calibrate and validate such methods. Currently, for images of real lesions the only accepted method to establish their size is based on annotations performed by expert radiologists.

In 2000 the National Institutes of Health launched a cooperative effort, known as the Lung Image Database Consortium [12], to construct a set of annotated lung images, especially low-dose helical CT scans of adults screened for lung cancer, and related technical and clinical data, for the development, the testing, and the evaluation of different computer-aided cancer screening and diagnosis technologies [13]. The LIDC developed a pulmonary nodule

documentation process [15,16] where expert radiologists marked the visible lesion boundary belonging to each lesion in all of the relevant axial images.

The LIDC annotation process did not require the expert radiologist to provide either uni- or bi-dimensional measurement, a technique commonly used in clinical practice. However, given the full boundary of the lesion as marked by the radiologist, we used computer algorithms to apply the rules from RECIST (uni-dimensional) and WHO (bi-dimensional) to provide estimates for the largest diameter and the largest perpendicular measures. We also computed volumetric-based measurements by processing the boundary documentations.

This paper addresses primarily the study of the absolute size of nodules, especially given its relevance to the CAD community, in two ways: by analyzing the variation between the expert markings and by comparing the fully three-dimensional volume measurements with the estimated RECIST, WHO and MS measurement methods, which involve a single two-dimensional image.

2 Materials and Methods

The evaluation of the impact of different size metrics was carried out on 265 documented whole-lung CT scans, of which 197 had nodules documented with radiologists' boundaries. All of the 197 scans were acquired from multi-detector row CT scanners with pixel size ranging from 0.508 to 0.946 mm (average 0.66 mm) and an axial slice thickness ranging from 0.625 to 3.000 mm (average 1.7 mm, median 1.8 mm). The tube current ranged from 40 to 582 mA (average 177.5 mA, median 160 mA), tube voltage range was for more than half cases 120kVp with the remaining ones having voltages equal to 130kVp (8), 135kVp (23), and 140 kVp (28). All the processing was performed on anonymized data, devoid of any identifying information in accordance to the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [14,15], that was provided by the LIDC institutions after approval by their respective IRBs.

As per the LIDC process model [15,16], each scan was assessed by 4 experienced thoracic radiologists, using calibrated monitors with magnification capabilities and detailed reading rules (e.g. initial window and level setting of 1500 HU and −500 HU) as described in [16]. Within those rules, the radiologists were instructed to mark the entire boundary, in all the relevant axial scan images, of all the nodules they estimated to be greater than 3mm in diameter (for smaller nodules, which were not included in this study, just the central location was requested). The outer boundary was chosen to be made of those pixels that were just outside the region of the nodule. As some nodules have cavities or holes in them, radiologists could also draw inner boundaries when they wanted to express the fact that a portion of the sorrounding region did not belong to the actual nodule. All the markings are stored as families of sets of boundary pixels located in the axial image planes. A typical marking on a single image is illustrated in Figure 1, where the radiologist drawn boundaries are superimposed to the CT scan image. Figure 2 shows in more detail the LIDC rules regarding the definition of the lesion region. Here, the original scan image is on the left while the marked image is on the right with the marker's boundary points shown in white and the region designating the marked lesion in black. The National Cancer Imaging Archive (NCIA) repository of the National Cancer Institute [17] makes available the CT scans that have been fully documented by the LIDC together with XML documentation files that contain the boundary points chosen by all of the expert radiologists.

All the scans and the XML files that were currently available as of this writing were imported and parsed to extract the radiologists' markings represented by outline information. For each nodule, the markings for each radiologist were converted, according to the LIDC process rules

(e.g. the black inner region in Figure 2b), into 3D binary occupacy images on which size measurements could be made. For example, a nodule that had been marked by just three of the four radiologists would have three corresponding 3D binary images, one for each radiologist.

The 3D images were used to compute the key values for the metrics under investigation: nodule volume, largest diameter and its largest perpendicular. The total lesion volume is estimated by counting the number of nodule pixels in each of the image slices and then multiplying their sum by the voxel volume [18]; this method is frequently used in CAD tools. Pixels belonging to the excluded inner regions do not belong to the nodule region and therefore are not counted were computing the nodule volume. The largest diameter is determined as the maximum diameter, i.e. the largest rectilinear segment that completely lies within the nodule region (see Figure 3), among all the axial-planar subsets of the nodule; this measure is similar to RECIST [2]. The computer estimation of the RECIST measurement is illustrated in Figure 3a. The solid line is the largest diameter that can be placed in any axial image within the marked boundary. When lesions have cavities or holes in them, as it is possible in the LIDC database, additional care must be taken to estimated the largest diameter. As a hypothetical example, if the radiologist had marked the pixels within the nodule shown white with an x in Figure 3b, this would imply a hole for that region that the radiologist wished to exclude. For the RECIST criterion the diameter must be within the lesion and not include the cavity; therefore, the diameter shown by the solid line would be the newly computer-determined RECIST measurement for that lesion.

For the WHO and MS measurements the largest perpendicular is also computed by considering every diameter pixel to determine the largest sum of the two half-perpendiculars stemming from that pixel. For the WHO measurement the lesion area a is computed as $\frac{\pi}{4} lw$ whereas, according to the MS equation [5], the lesion volume v is computed as being equal to $\frac{\pi}{6} l * w^2$; in both cases, l is the largest diameter length and w is the length of its largest perpendicular.

For our statistical evaluation, each measure is made equivalent and directly comparable to the others by expressing its value in terms of the diameter of an equivalent sphere or circle according to the measure dimension. For any given volume estimate the equivalent diameter d of a sphere with the same volume is calculated using $2\sqrt[3]{\frac{3v}{4\pi}}$, where the volume of the equivalent sphere is $v = \frac{4}{3}\pi\left(\frac{d}{2}\right)^3$. For any given area estimate the equivalent diameter d of a circle with the same area is calculated using $2\sqrt{\frac{a}{\pi}}$, where the area of the equivalent circle is $a = \pi\left(\frac{d}{2}\right)^2$. Hence, for the bi-dimensional and MS metrics, the measured size is equivalent to a diameter of $\sqrt{lw}$ and $\sqrt[3]{l * w^2}$ respectively, where l is the largest diameter length and w is the length of its largest perpendicular.

For the inter-reader variation, we considered only those nodules that had markings by all four radiologists. For those nodules we computed means and standard deviations based on radiologists' equivalent measurements. We also computed the smoothed estimators of the variation as a function of the nodule size.

For the metric comparison when multiple markings were available for a nodule, the median value of each size metric from these markings was used to represent the size for that nodule. From a preliminary analysis we observed that the measurement distributions were non-symmentrical and possibly non-unimodal. Therefore, we chose to use a non-parametric method to estimate the probability density functions because they are more flexible and do not assume

a known functional form; in particular, we selected kernel estimators that work by smoothing out the contribution of each observed data point over a local neighbourhood of that data point [19]. We determined the kernel estimates of the conditional distribution of the volume measurement given the uni-dimensional, bi-dimensional, and MS sizes. From the estimates of the conditional distribution, estimates of conditional highest density regions (HDR) [20] were computed. HDRs are subsets of the measurement values for which all the points inside a region have a higher probability density than all the points outside that region and they provide tight confidence intervals with superior coverage properties than their alternatives. From these we computed the 0.95 and 0.99 probability coverage regions to represent our comparative conditional distributions.

3 Results

The whole dataset contained 522 lesions for which boundaries were marked. Four nodules had diameters derived from their three-dimensional measures greater than 30 mm, (33, 36, 48, and 68 mm respectively) and were not considered in the analysis because they fell outside the LIDC nodule size range [16]. Of these 518 lesions only 127 were considered to be nodules greater than 3 mm by all four radiologists and, therefore, had four sets of boundary markings available.

The between reader variation was analyzed for 127 nodules with four readings each and three-dimensional sizes ranging from 3.36 to 23.8 mm (average 9.01 mm, median 7.37 mm). Figure 4 shows the smoothed estimators for the four metrics (two data points fell outside the plot area for the uni-dimensional metric with values of (25.7,7.3) and (36.8,3.2), and one for the MS metric with values of (16.3,5.1)). For the three-dimensional metric 95% of estimated standard deviations were within the following range [0.49, 1.25]. For the uni-dimensional, bi-dimensional, and MS size metric most of the standard deviations (95%) were within the ranges [0.67, 2.55], [0.78, 2.11] and [0.96, 2.69] respectively. For the three-dimensional derived diameters, 2 nodules (1.6%) were not within two standard deviations while 15 nodules (11.8%), 13 nodules (10.2%), and 25 nodules (19.7%) had a standard deviation greater than two for the uni-dimensional, bi-dimensional, and MS metrics respectively. As an example, for the three-, uni-, bi-dimensional and MS size measurements of 10 mm, the estimated standard deviations were 0.85, 1.16, 1.18 and 1.39 respectively. If the size under consideration is 15 mm, then the estimated standard deviations are 1.09, 1.59, 1.54, and 1.79 respectively.

The full set of 518 nodules we considered for the metric comparison had nodule sizes, based on the three-dimensional metric, ranging from 2.03 to 29.4 mm (average 7.05 mm, median 5.71 mm); Figure 5 shows their size distribution. 452 nodules out of 518 (87.3%) met the recommendations of RECIST [2]. The remaining 66 nodules (average three-dimensional metric size 4.42 mm, median 4.60 mm) were technically too small with respect to slice thickness to meet the RECIST recommendations for measurement ([2], Appendix 1); in detail 30 were on 16 scans with a slice thickness of 3.00 mm, 24 were on 15 scans with a slice thickness of 2.50 mm, and the remaining 12 on 11 scans with a slice thickness between 1.25 and 2.00 mm. The RECIST rule to skip inner holes was applied in 17 cases with an average relative decrease in length of 5.5% (median 4.5%) with respect to maximal segments that ignore the rule. Figure 6 shows the case where the change was the largest (4.11 mm, 16.5%) because of the inner region with the light boundary was marked as not being part of the nodule.

Figure 7 shows HDRs for the three-dimensional metric with 95 and 99 percent coverage conditional on the uni-dimensional, bi-dimensional, and MS size measurements. The conditional medians are marked with dots. For example, if a radiologist reports a uni-dimensional size measurement of 10 mm, then the regions of probability coverage 0.95 and 0.99 with the smallest extent for the the three-dimensional metric measurement are the intervals [4.16, 11.48] and [3.60, 12.21] respectively. If the 10mm measure was from a bi-dimensional

measurement, then the region intervals would be [5.04, 12.77] and [3.65, 13.29] respectively. In the case of the MS metric, they would be [5.33, 11.61] and [4.24, 12.69] respectively. For comparison, if the 10mm measure is from the bi-dimensional metric, the regions of probability coverage 0.95 and 0.99 with the smallest extent for the MS metric are [10.62, 12.61] and [10.54, 12.63] respectively, which respectively represent a 74% and 78% relative decrease in the range of the measure estimation.

4 Discussion

The evaluation we performed on a set of 265 documented whole-lung scans focused on two aspects: (a) the inter-reader variability and its relationship to the analyzed metrics and (b) the level of agreement between the analyzed metrics.

Previous studies [21,22] evaluated inter- and intra-observer variability on single image measures, either mono- or bi-dimensional, and they found considerable variability in those measures. Meyer et al. [23] took advantage of the boundaries drawn by six radiologists around 23 lung nodules to compute lung nodule volumes showing that radiologists' subjectivity is a major source of variability.

Our analysis of 127 pulmonary nodules found that reader subjectivity on boundary locations may propagate into a very large inter-observer variation of the size estimates. It is clear, as shown by Figure 4, that the uni-dimensional, bi-dimensional, and MS size metrics have a slightly higher inter-observer variation than the three-dimensional metric. Additionally, the smoothed estimate of the three-dimensional metric variation starts lower and increases more slowly than the variation estimates of the uni-dimensional, bi-dimensional and MS metrics. One reason for this high variability is that, when the nodule has a complex shape, each radiologist may mark some boundary pixels as belonging or not belonging to the nodule region. This in turn is reflected in diameters and perpendiculars that can show significant differences in length as shown in Figure 8.

As far as the comparison between metrics is concerned, a previous study by Van Hoe et al. in 1997 [24] compared five different measurement methods for change in lesion size between scans on CT images of liver metastases. The authors concluded that the three-dimensional methods were a viable alternative to two dimensional methods. Other studies on size change for treatment response assessment in lung cancer have compared uni-dimensional and automated three-dimensional measures [21], uni-, bi- and three-dimensional measures [25, 26], and manual bi-dimensional measures with an automated contour technique [27]. The studies that involved volume measurements concluded that there was poor agreement between volumetric-based measures and single image based methods.

Our analysis of 518 nodules, marked by at least one radiologist, showed that the differences between the single image measurements and the three-dimensional measurements are very large. The 0.95 and 0.99 probability coverage regions are very wide and their projections on the ordinate axis have large overlaps; moreover the intervals are not centered aroud the marginal values, implying an underlying bias or a size-dependent scaling factor between the measures as well.

Poor agreement between the three-dimensional and two-dimensional methods is commonly seen when the nodule does not conform to the approximately spherical or ellipsoidal assumptions that underlie the one and two dimensional measurements respectively. Examples of this are shown in Figures 9 and 10. In Figure 9 the largest extent of the lesion is aligned with the axial dimension, hence the two-dimensional measurements under-estimates the three-dimensional derived diameter while, in Figure 10, the largest extent of the nodule is perpendicular to the axial dimension, hence the two-dimensional measurements over-estimates

the three-dimensional derived diameter. In general, for the uni-dimensional metric we anticipated an over estimation of the three-dimensional value since it is a measure of maximal extent rather than mean extent in two dimensions. In the context of CAD algorithm evaluation the selection of the size metric may have an important impact. For example, consider that a CAD system is designed to detect pulmonary nodules larger than 6 mm; to evaluate this system we select from the documented database all nodules larger than 6 mm as the set of true positives. For the LIDC data set if we use the 6 mm size limit the number of nodules selected from the 518 total nodules is as follows: three-dimensional 228, uni-dimensional 310, bi-dimensional 197, and MS 242.

A further problem is the difference between the selected subsets. For example, if we compare the selections for the three-dimensional and bi-dimensional metrics, we find that 45 nodules in the three-dimensional metric list are not in the bi-dimensional metric list and 14 nodules that are in the bi-dimensional metric list are not in the three-dimensional metric list. The consequence is that the measured system sensitivity is reduced if different metrics for size cutoff are used for test set selection and CAD system implementation. Consider the perfect CAD system that will detect every nodule marked by the LIDC and that it uses the three-dimensional metric for a 6 mm minimum size cutoff. If we test this system selecting from the LIDC cases also using the three-dimensional metric then all the nodules will be correctly identified for 100% sensitivity. However, if we test the same system with the selection by the bi-dimensional metric criterion then, of the 197 nodules, 14 will not be identified because they are considered to less than 6 mm; therefore, the measured sensitivity will be 183/197 or 95%. In addition 45 nodules only in the three-dimensional list would be considered as false positives. If the uni-dimensional metric had been used for selecting the test cases, then the measured sensitivity would be further reduced to 223/310 or 72%.

The selection of data subsets with size limits can only be directly compared if the same size metric is used in both cases. The LIDC plans to include a single size measure for each nodule in the database to facilitate the selection of unique repeatable size limited nodule subsets. This metric is not intended as a gold standard for nodule size. Further, when evaluating the performance of CAD detection algorithms any difference between the data selection metric and the algorithm size selection criterion needs to be considered.

Acknowledgements

This research was funded in part by the National Institutes of Health, National Cancer Institute, Cancer Imaging Program by the following grants: R33 CA101110, R01 CA078905, 1U01 CA 091085, 1U01 CA 091090, 1U01 CA 091099, 1U01 CA 091100, and 1U01 CA 091103.

References

1. WHO handbook for reporting results of cancer treatment, World Health Organization Offset Publication No. 48 (1979).
2. Therasse P, Arbuck S, Eisenhauer E, Wanders J, Kaplan R, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom A, Christian M, Gwyther S. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 2000;92(3):205–216. [PubMed: 10655437]
3. Hasegawa M, Sone S, Takashima S, Li F, Yang ZG, Maruyama Y, Watanabe T. Growth rate of small lung cancers detected on mass CT screening. British Journal of Radiology 2000;73:1252–1259. [PubMed: 11205667]
4. Lindell RM, Hartman TE, Swensen SJ, Jett JR, Midthun DE, Tazelaar HD, Mandrekar JN. Five-year lung cancer screening experience: CT appearance, growth rate, location, and histologic features of 61 lung cancers. Radiology 2007;242(2):555–562. [PubMed: 17255425]
5. Usuda K, Saito Y, Sagawa M, Sato M, Kanma K, Takahashi S, Endo C, Chen Y, Sakurada A, Fujimura S. Tumor doubling time and prognostic assessment of patients with primary lung cancer. Cancer 1994;74(8):2239–2244. [PubMed: 7922975]

6. Ko JP, Rusinek H, Jacobs EL, Babb JS, Betke M, McGuinness G, Naidich DP. Small pulmonary nodules: Volume measurement at chest CT –phantom study. Radiology 2003;228(3):864–870. [PubMed: 12954901]
7. Kuhnigk, J-M.; Dicken, V.; Bornemann, L.; Wormanns, D.; Krass, S.; Peitgen, H-O. Lecture Notes in Computer Science, Vol. 3217, Medical Image Computing and Computer-Assisted Intervention. Springer-Verlag GmbH; 2004. Fast automated segmentation and reproducible volumetry of pulmonary metastases in CT-scans for therapy monitoring; p. 933-941.
8. Okada K, Comaniciu D, Krishnan A. Robust anisotropic gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT. IEEE Transactions on Medical Imaging 2005;24(3):409–423. [PubMed: 15754991]
9. Goodman LR, Gulsun M, Washington L, Nagy PG, Piacsek KL. Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. American Journal of Roentgenology 2006;186:989–994. [PubMed: 16554568]
10. Revel MP, Merlin A, Peyrard S, Triki R, Couchon S, Chatellier G, Frija G. Software volumetric evaluation of doubling times for differentiating benign versus malignant pulmonary nodules. American Journal of Roentgenology 2006;187:135–142. [PubMed: 16794167]
11. Reeves A, Chan A, Yankelevitz D, Henschke C, Kressler B, Kostis W. On measuring the change in size of pulmonary nodules. IEEE Transactions on Medical Imaging 2006;25(4):435–450. [PubMed: 16608059]
12. National Institutes of Health. Lung image database resource for imaging research. [accessed August 28, 2007 (April 2000)]. http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-01–001.html
13. National Cancer Institute. Lung Imaging Database Consortium (LIDC). [accessed August 28, 2007]. http://imaging.cancer.gov/programsandresources/InformationSystems/LIDC
14. Department of Health and Human Services. Unofficial Version of HIPAA Administrative Simplification Regulation Text, 45 CFR Parts 160, 162, and 164, 2006. [accessed August 28, 2007]. http://www.hhs.gov/ocr/AdminSimpRegText.pdf
15. McNitt-Gray MF, Armato SG III, Meyer CR, Reeves AP, McLennan G, Pais R, Freymann J, Brown MS, Engelmann RM, Bland PH, Laderach GE, Piker C, Guo J, Towfic Z, Qing DP, Yankelevitz DF, Aberle DR, van Beek EJR, MacMahon H, Kazerooni EA, Croft BY, Clarke L. The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. Academic Radiology. in press
16. Armato SG III, McLennan G, McNitt-Gray MF, Meyer CR, Yankelevitz D, Aberle DR, Henschke CI, Hoffman EA, Kazerooni E, MacMahon H, Reeves AP, Croft BY, Clarke LP. the Lung Image Database Consortium Research Group. Lung image database consortium: developing a resource for the medical imaging research community. Radiology 2004;232(3):739–748. [PubMed: 15333795]
17. National Cancer Institute. National Cancer Imaging Archive. [accessed August 28, 2007]. https://imaging.nci.nih.gov/ncia/
18. Breiman RS, Beck JW, Korobkin M, Glenny R, Akwari OE, Heaston DK, Moore AV, Ram PC. Volume determinations using computed tomography. American Journal of Roentgenology 1982;138 (2):329–333. [PubMed: 6976739]
19. Hyndman RJ, Bashtannyk DM, Grunwald GK. Estimating and visualizing conditional densities. Journal of Computational and Graphical Statistics 1996;5(4):315–336.
20. Hyndman RJ. Computing and graphing highest density regions. The American Statistician 1996;50 (2):120–126.
21. Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. European Radiology 2006;16(4):781–790. [PubMed: 16331462]
22. Bogot NR, Kazerooni EA, Kelly AM, Quint LE, Desjardins B, Nan B. Interobserver and intraobserver variability in the assessment of pulmonary nodule size on CT using film and computer display methods. Academic Radiology 2005;12(8):948–956. [PubMed: 16087090]
23. Meyer CR, Johnson TD, McLennan G, Aberle DR, Kazerooni EA, MacMahon H, Mullan BF, Yankelevitz DF, van Beek EJR, Armato SG III, McNitt-Gray MF, Reeves AP, Gur D, Henschke CI, Hoffman EA, Bland PH, Laderach G, Pais R, Qing D, Piker C, Guo J, Starkey A, Max D, Croft BY,

Clarke LP. Evaluation of lung MDCT nodule annotation across radiologists and methods. Academic Radiology 2006;13:1254–1265. [PubMed: 16979075]

24. Van Hoe L, Van Cutsem E, Vergote I, Baert AL, Bellon E, Dupont P. Size quantification of liver metastases in patients undergoing cancer treatment: reproducibility of one-, two-, and three-dimensional measurements determined with spiral CT. Radiology 1997;202:671–675. [PubMed: 9051014]
25. Tran LN, Brown MS, Goldin JG, Yan X, Pais RC, McNitt-Gray MF, Gjertson D, Rogers SR, Aberle DR. Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography. Acad Radiol 2004;11(12):1355–1360. [PubMed: 15596373]
26. Jennings SG, Winer-Muram HT, Tarver RD, Farber MO. Lung tumor growth: Assessment with CT – comparison of diameter and cross-sectional area with volume measurements. Radiology 2004;231 (3):866–871. [PubMed: 15163822]
27. Schwartz LH, Ginsberg MS, DeCorato D, Rothenberg LN, Einstein S, Kijewski P, Panicek DM. Evaluation of tumor measurements in oncology: Use of film-based and electronic techniques. Journal of Clinical Oncology 2000;18(10):2179–2184. [PubMed: 10811683]

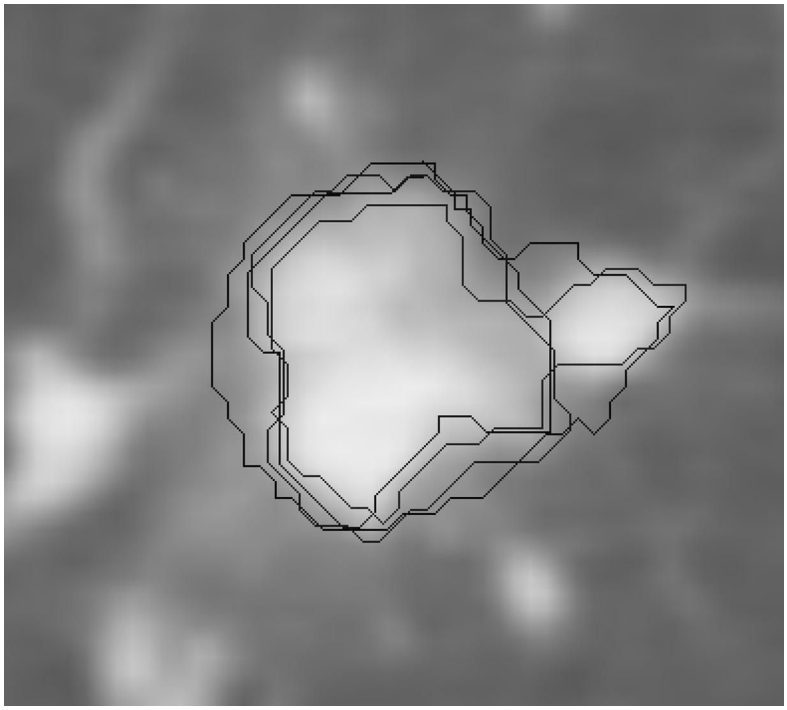


Fig. 1.
An example of a single image section of the markings provided by the LIDC database.

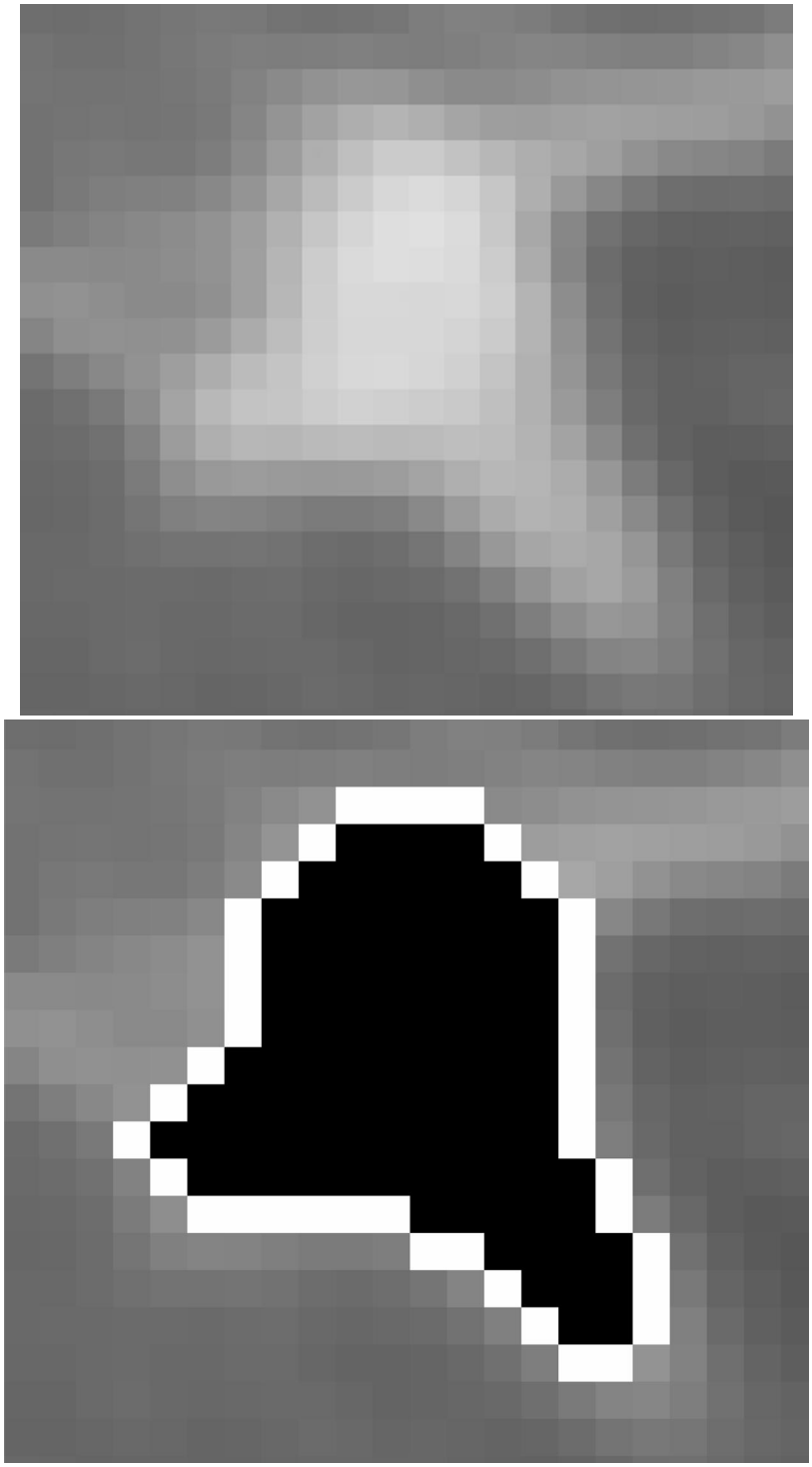


Fig. 2.
An example of the LIDC rules in documenting nodules. On the left (a), the original image data is presented. On the right (b), the white boundary shows the actual boundary drawn by the radiologist that encloses the black inner region belonging to the nodule.

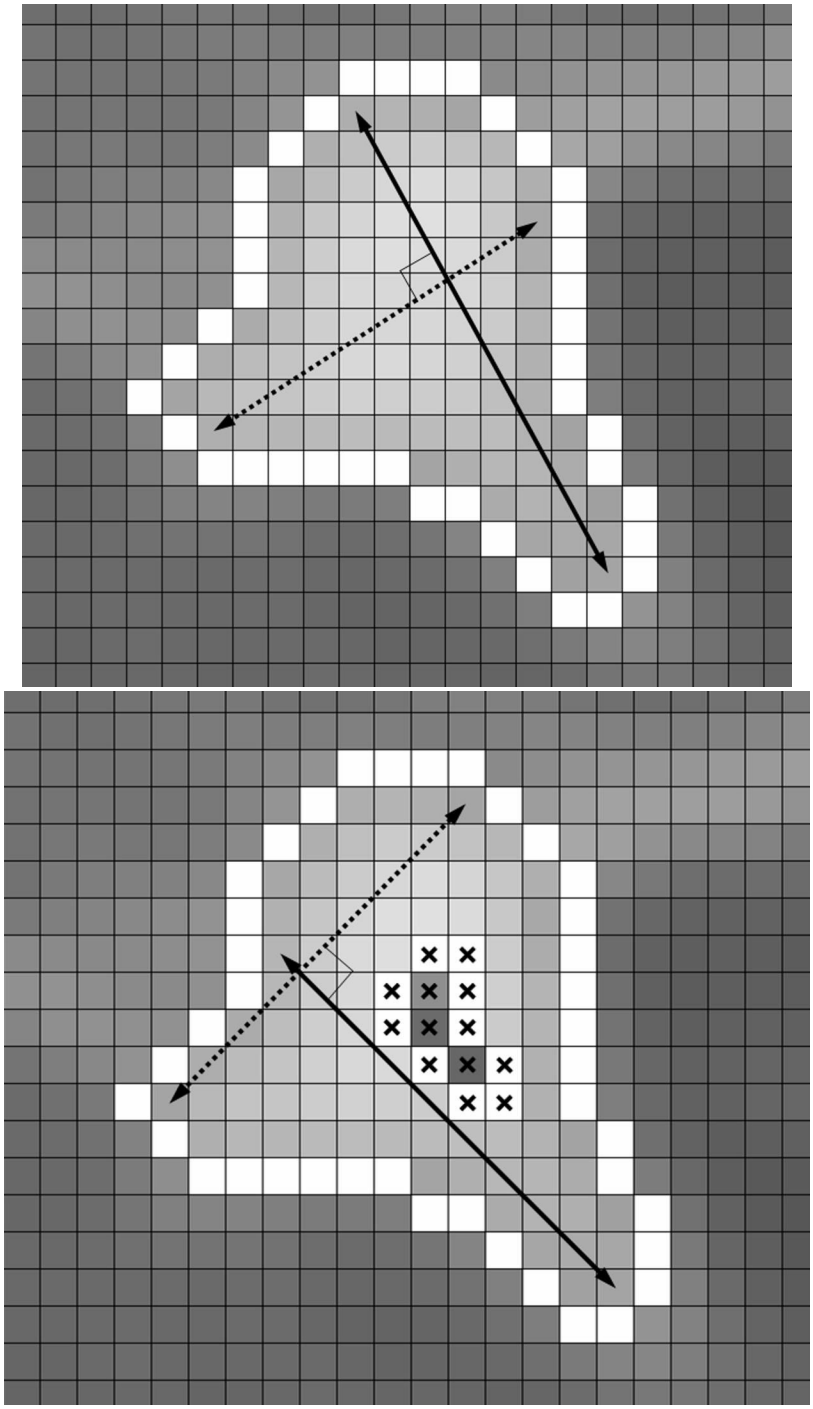
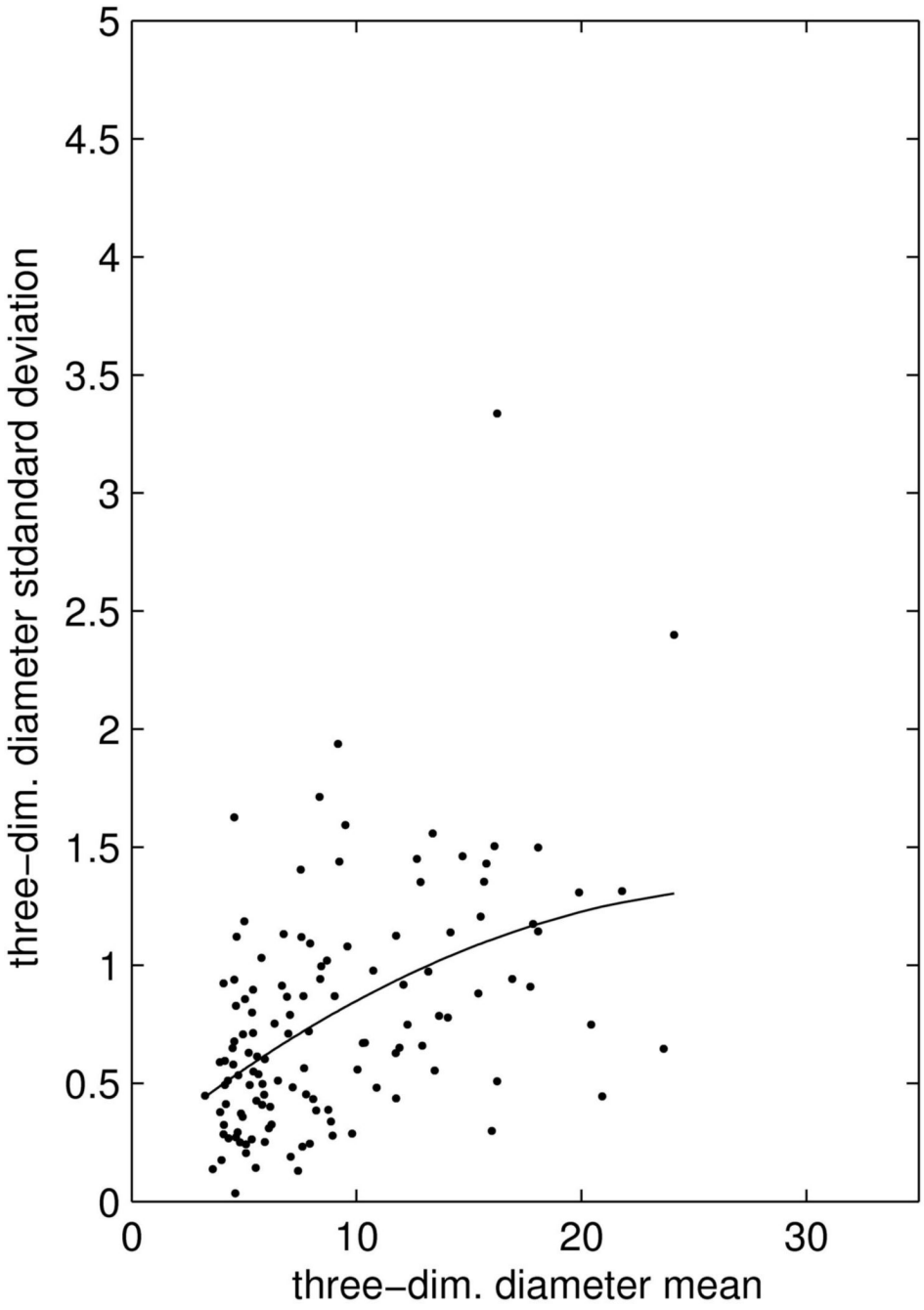
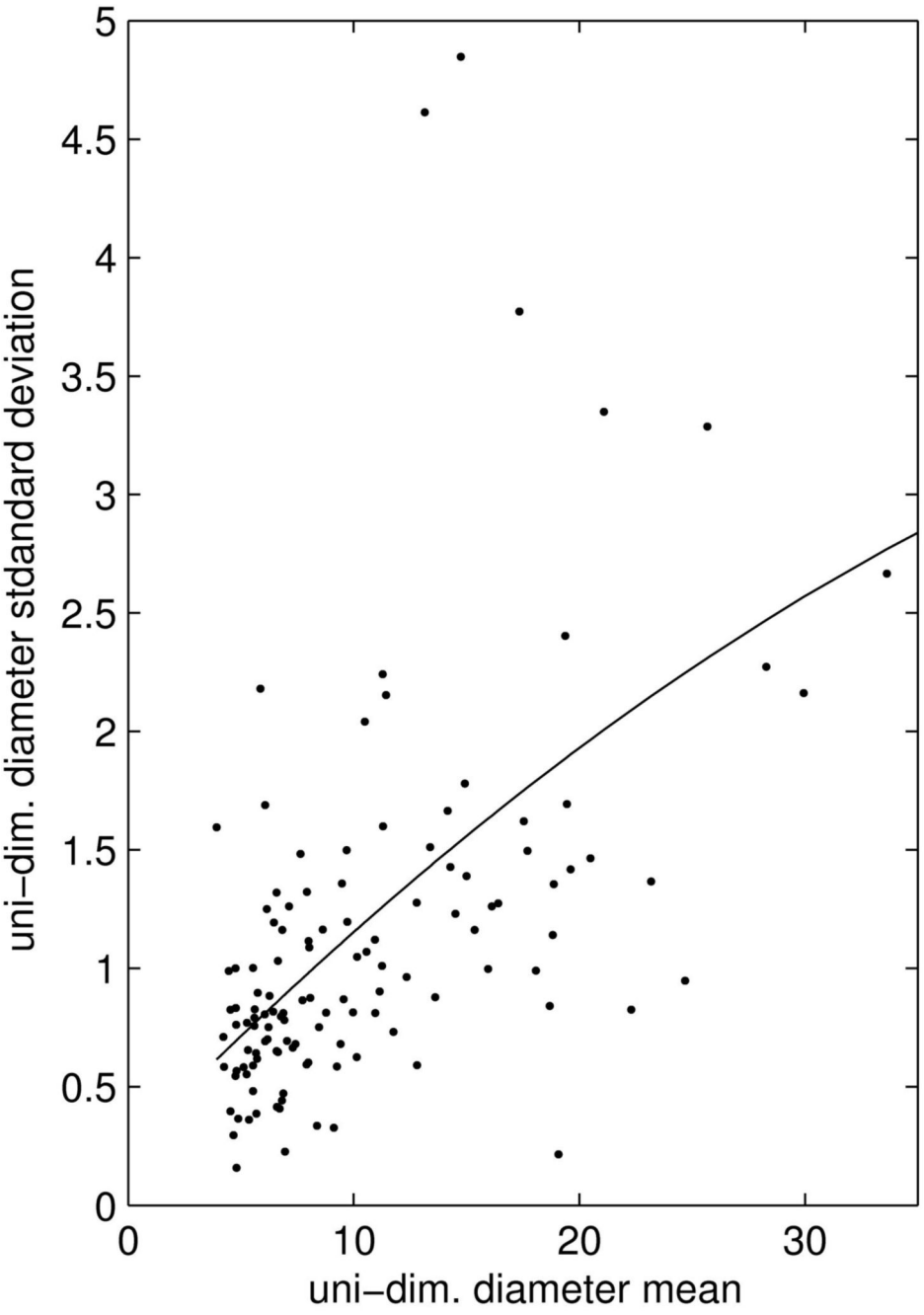
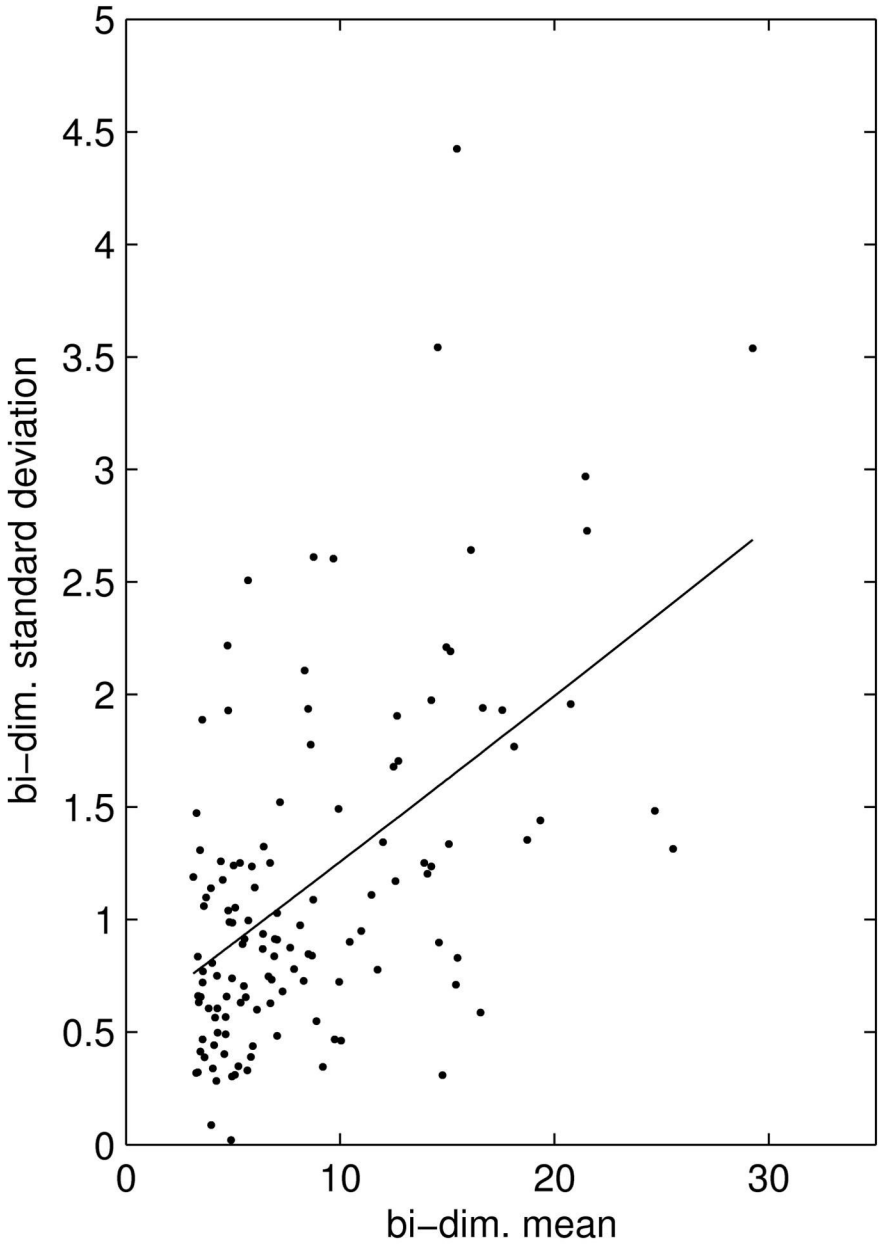


Fig. 3.

This figure, on the left (a), describes graphically how the diameter and its largest perpendicular are computed as surrogates of radiologist actions. On the right (b), if the sub-region with the pixels marked with a cross were to be hypothetically removed from the actual nodule region, then the previous diameter would not be valid any longer and the new diameter with the relative largest perpendicular would have to be determined. The three-dimensional metric size would be affected, too, being computed on the decreased nodule volume.







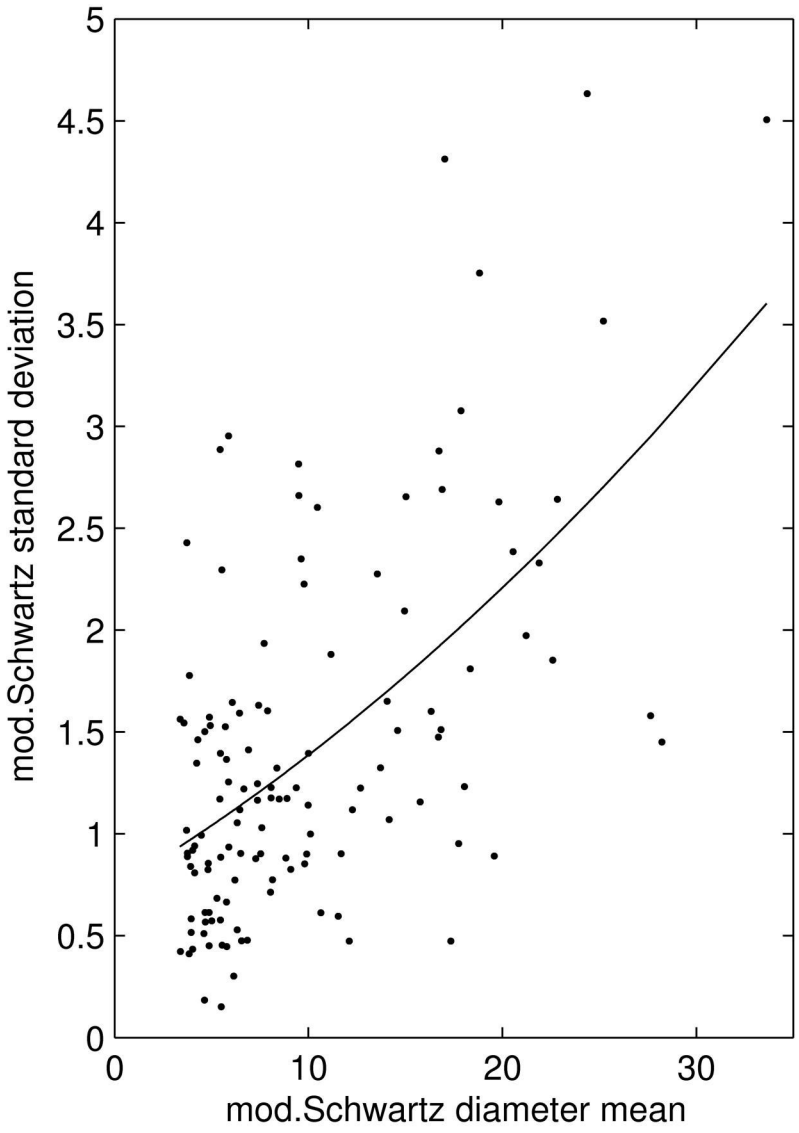


Fig. 4.

Scatter plot of the standard deviation versus means of four experts' measurements along with a non-parametric regression curve for three-dimensional (a), uni-dimensional (b), bi-dimensional (c), and MS (d) size estimates.

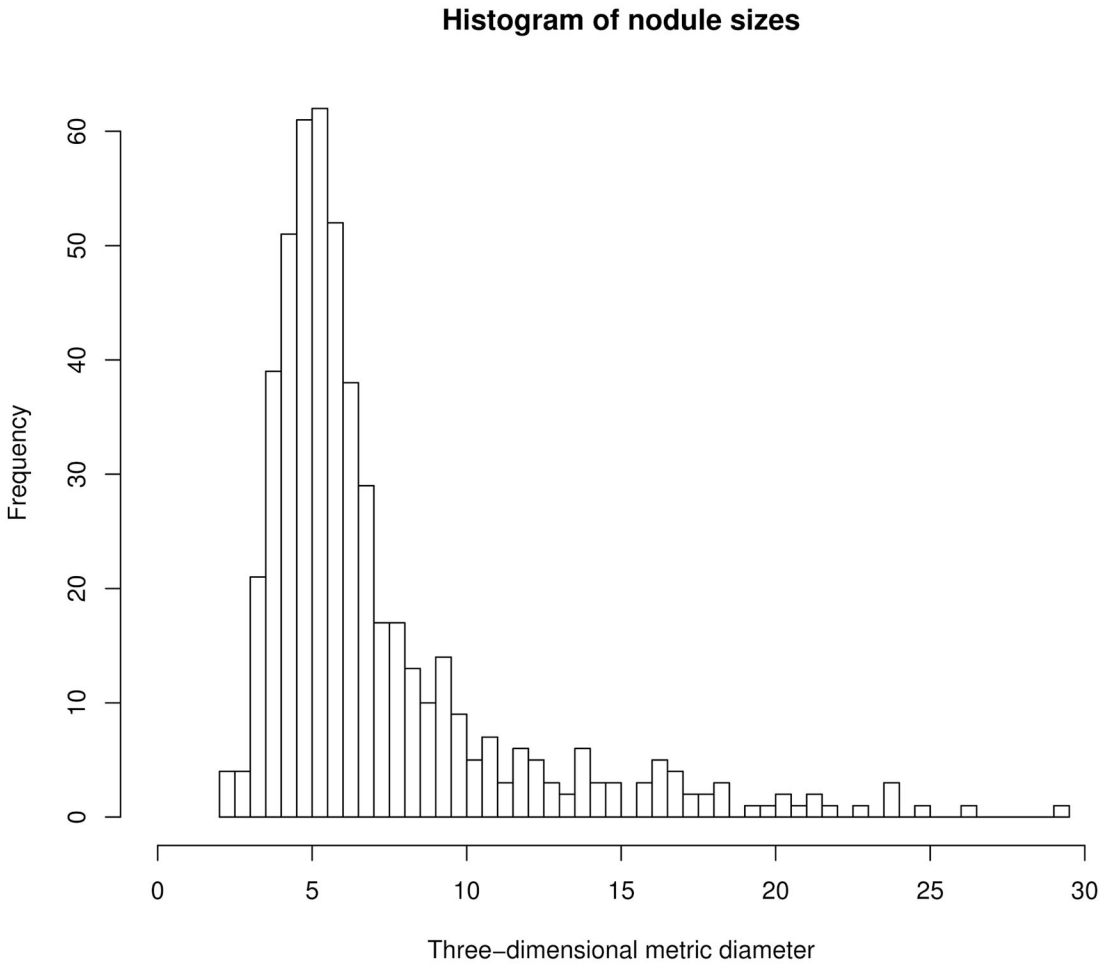
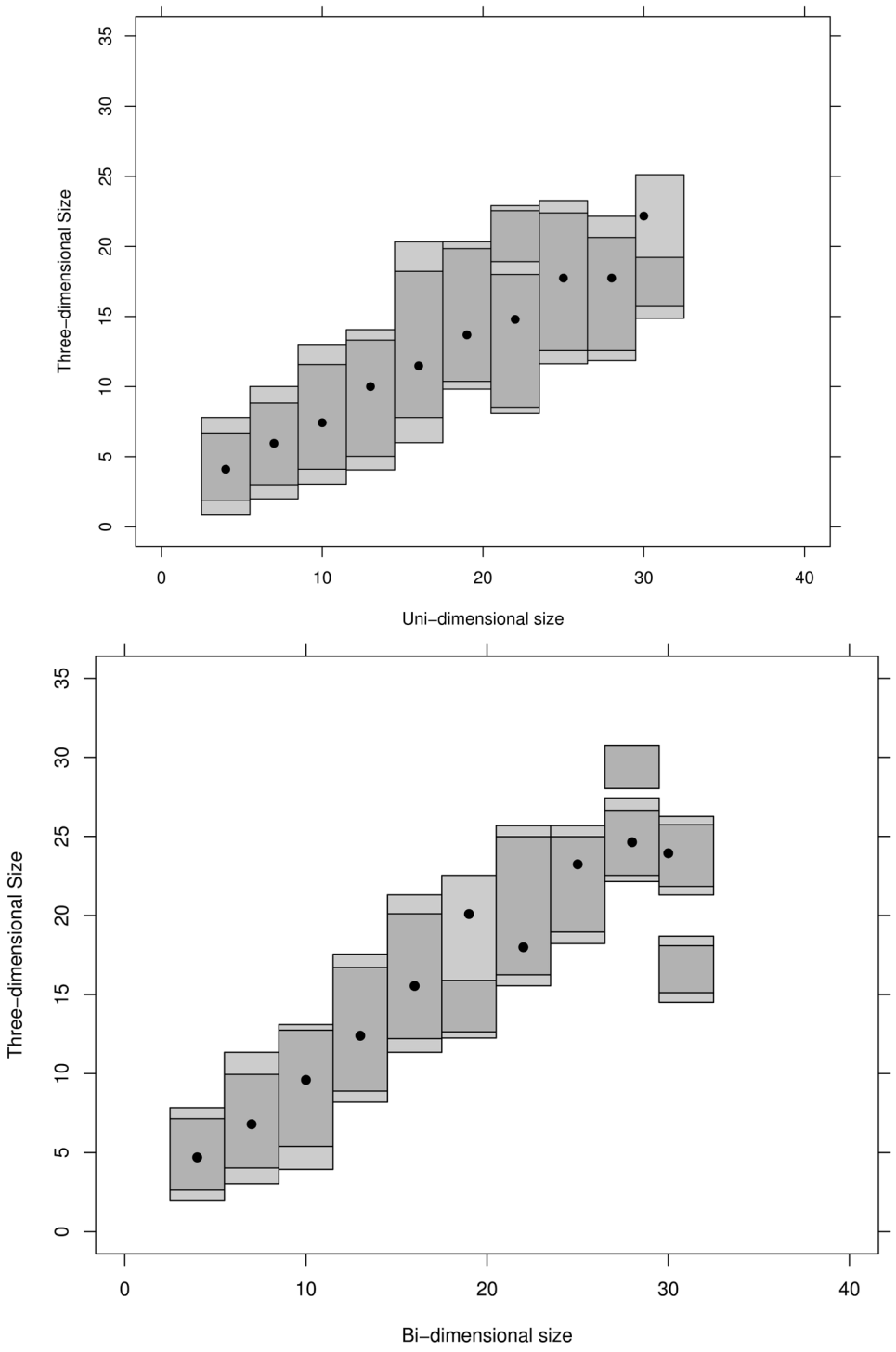


Fig. 5.
The size distribution (according to the three-dimensional metric) of the full set of 518 nodules.

Fig. 6.
A nodule with an inner region marked by a light boundary. As the inner region and its boundary are not part of the nodule, the depicted segment cannot be considered a diameter by the RECIST rules.



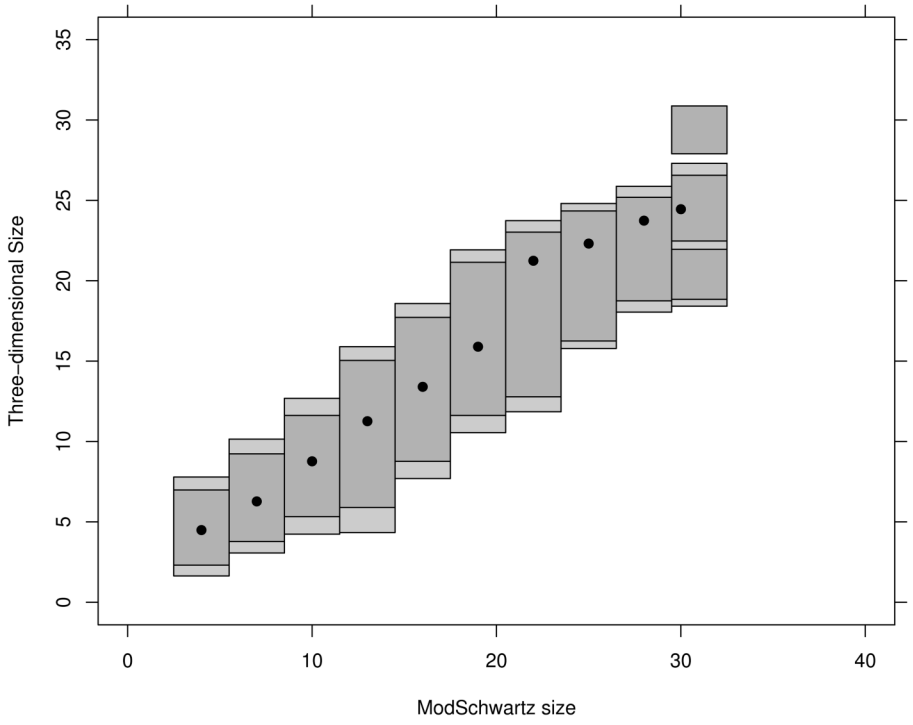
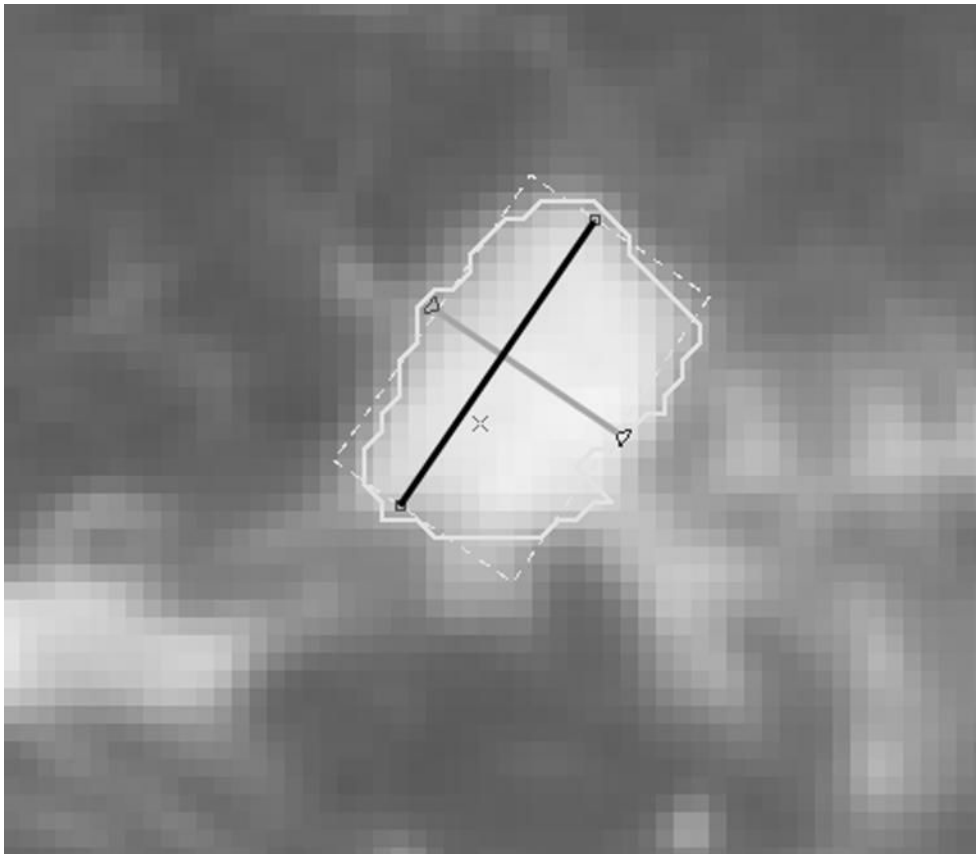
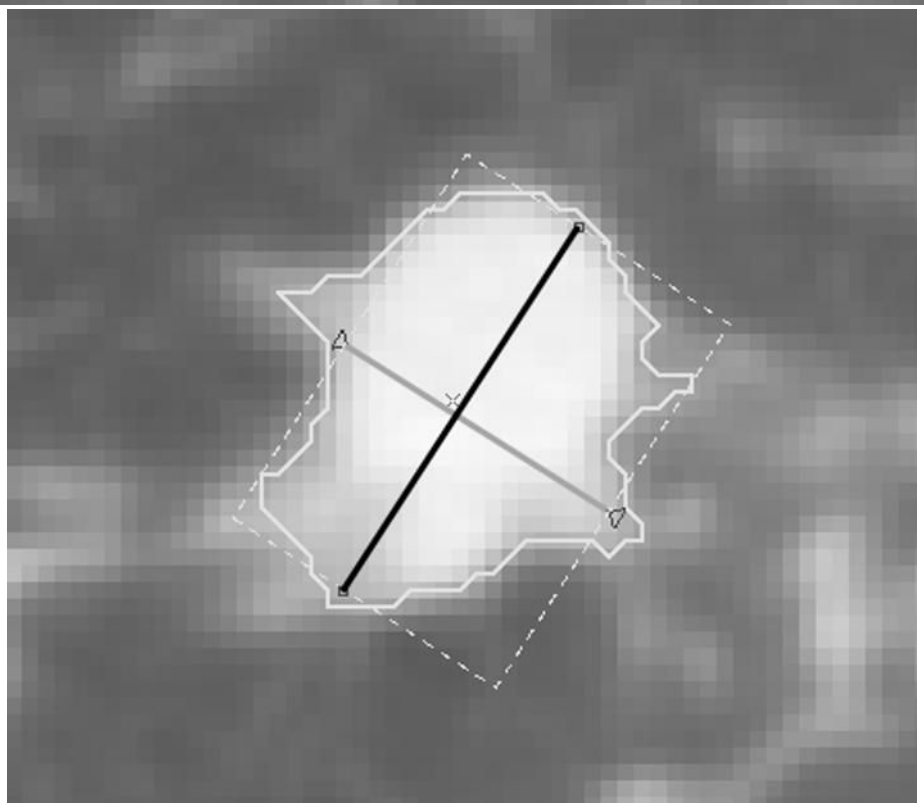
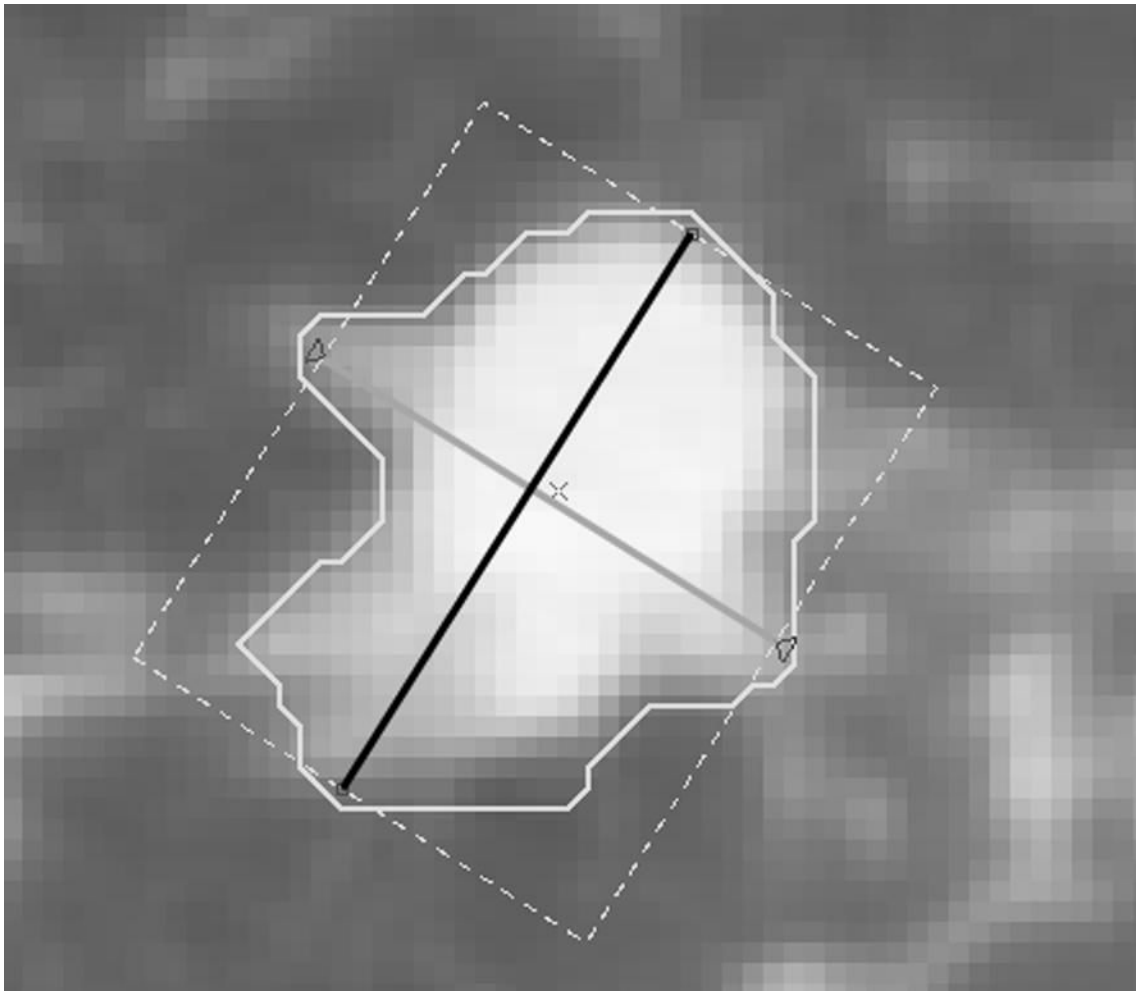


Fig. 7.

95% and 99% HDRs for the three-dimensional metric size estimate conditional on the uni-dimensional metric (a), on the bi-dimensional metric (b), and on the MS metric (c).







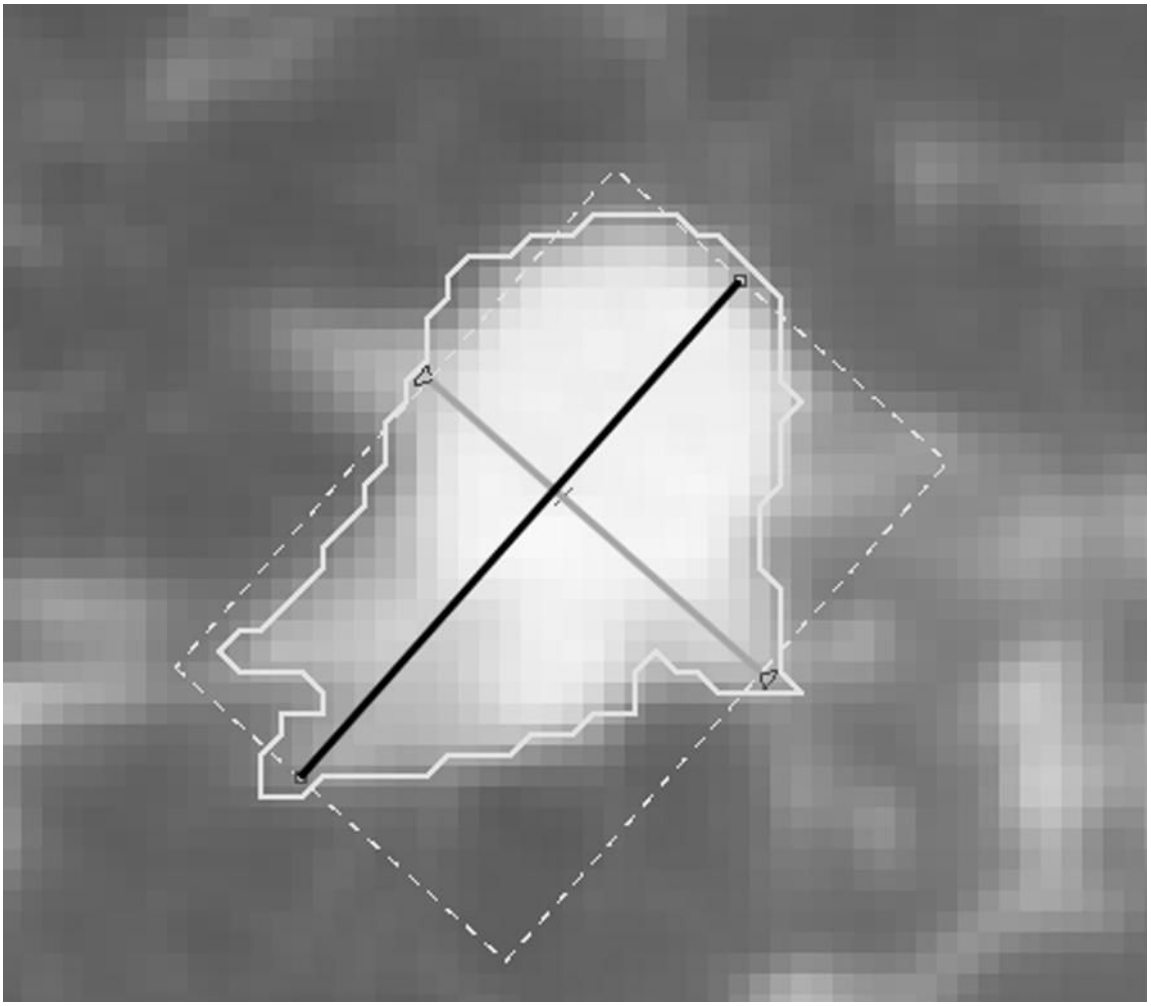


Fig. 8.

An example of variability among radiologists. Each image shows the slice where the largest diameter (dark line) and largest perpendicular (gray line) were determined according to the markings provided by each of the four radiologists (a-d). The first image (a) is on a different slice than the other three (b-d); this is possible since each slice selected for measurement is based on a radiologist's individual marking.

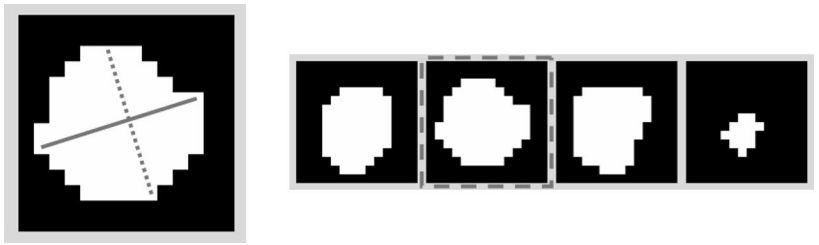


Fig. 9.
A selected case where the three-dimensional size (10.0 mm) is greater than the uni-dimensional (8.3 mm), bi-dimensional (8.0 mm), and MS (7.9 mm) sizes. The tiled frames on the right hand of the figure show all the nodule regions, in consecutive axial slices, used to compute the three-dimensional metric measure. The frame with dashed boundary is enlarged on the left hand of the figure to show the largest diameter (solid line) and its largest perpendicular (dotted line).

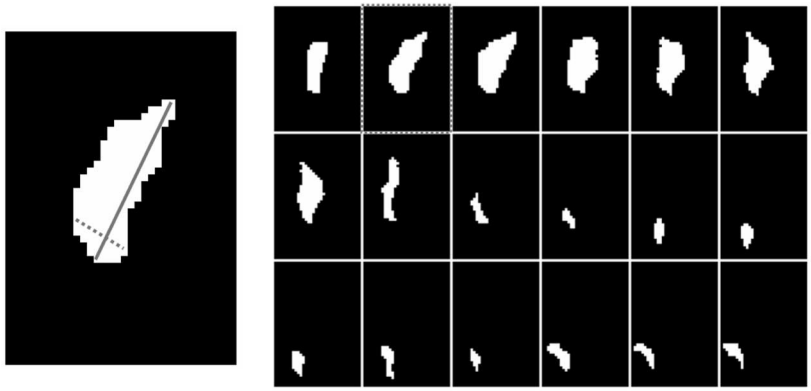


Fig. 10.
A selected case where the three-dimensional size (10.6 mm) is smaller than the uni-dimensional (21.7 mm), bi-dimensional (14.1 mm), and MS (12.2 mm) sizes. The tiled frames on the right hand of the figure show all the nodule regions, in consecutive axial slices, used to compute the three-dimensional metric measure. The frame with the dotted boundary is enlarged on the left hand of the figure to show the largest diameter (solid line) and its largest perpendicular (dotted line).